DOCUMENT RESUME

ED 471 544 TM 034 646

AUTHOR	Weems, Gail H.; Onwuegbuzie, Anthony J.; Lustig, Daniel
TITLE	Profiles of Respondents Who Respond Inconsistently to Positively- and Negatively-Worded Items on Rating Scales.
PUB DATE	2002-11-00
NOTE	33p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Chattanooga, TN, November 6-8, 2002).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC02 Plus Postage.
DESCRIPTORS	*Individual Characteristics; Likert Scales; *Profiles; *Rating Scales; *Responses; *Test Items
IDENTIFIERS	*Inconsistency

ABSTRACT

Many instruments, especially Likert-type scales, contain both positively and negatively worded items within the same scale (i.e., mixed item format). A major reason for this practice appears to be to discourage response sets from emerging. Using this format also helps the analyst detect response sets that occur in data sets, and thus eliminate them from subsequent analyses. Some psychometricians question the use of mixed item formats, positing that positively and negatively worded items within a scale are not measuring the same underlying trait. Limited evidence has suggested that certain individuals are more predisposed to providing differential response patterns when responding to a mixed item format scale. However, to date, only a few characteristics of these differentially responding individuals have been identified. This study sought to extend this research, analyzing responses to several scales using mixed item formats. Using a canonical correlation analysis, a sample of 158 students revealed a relationship between degree of differential response between positively and negatively worded items on three six-item measures of foreign language anxiety (i.e., input, processing, and output anxiety) and several dimensions of self-perception, study skills, and locus of control. Implications of all findings are discussed. (Contains 1 table and 61 references.) (Author/SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 471 544

Running Head: PROFILES OF RESPONDENTS

Profiles of Respondents Who Respond Inconsistently to Positively- and Negatively-Worded Items on Rating Scales

Gail H. Weems

The University of Memphis

Anthony J. Onwuegbuzie

Howard University

Daniel Lustig

The University of Memphis

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

G. Weems

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Correspondence should be addressed to Gail H. Weems, University of Memphis, 100 Ball Education Building, Memphis, TN 38152-3570, or E-Mail: (ghweems@memphis.edu).

TM034646

Paper Presented at the Annual Conference of the Mid-South Educational Research Association, Chattanooga, Tennessee 6-8 November 2002.

BEST COPY AVAILABLE

Abstract

Many instruments, especially Likert-type scales, contain both positively- and negatively-worded items within the same scale (i.e., mixed item format). A major reason for this practice appears to be to discourage response sets from emerging. Using this format also helps the analyst to detect response sets that occur in data sets, and thus eliminate them from subsequent analyses. However, some psychometricians seriously question the use of mixed item formats, positing that positively- and negatively-worded items within a scale are not measuring the same underlying trait. Limited evidence has suggested that certain individuals are more predisposed to providing differential response patterns when responding to a mixed item format scale. However, to date, only a few characteristics of these differential-responding individuals have been identified. Thus, the purpose of this present study was to extend this line of research. Specifically, the researchers analyzed responses to several scales utilizing mixed item formats. For example, using a canonical correlation analysis, a sample of 185 students revealed a relationship between degree of differential responses between positively- and negatively-worded items on three 6-item measures of foreign language anxiety (i.e., input, processing, and output anxiety) and several dimensions of self-perception, study skills, and locus of control. Implications of all findings are discussed.

Profiles of Respondents Who Respond Inconsistently to Positively- and Negatively-worded Items on Rating Scales

Negatively-worded items, those phrased in the semantically opposite direction of the majority (Barnette, 2000), are often posed on surveys in an attempt to diminish non-attending behaviors such as acquiescence, satisficing, and response set. Cronbach (1946, 1950) termed the tendency of participants simply to agree with survey items acquiescence. Later, Couch and Keniston (1960) labeled the predisposition toward the direction of item wording regardless of content "yea- or nay-saying." Krosnick, Narayan, and Smith (1996) attribute participants yea-saying to satsificing, or the proclivity to agree with an item due to the exertion of minimal cognitive effort. Irrespective of the cause, both behaviors may lead to participant responses that may not accurately communicate the trait or belief that the surveyor sought to measure. Measurement error on surveys also may result from, response set, or the tendency of participants to respond to general feelings about the survey topic rather than the specific item content. Therefore, negatively-worded items are often presented on surveys under the assumption that the items will encourage respondents to process items more carefully. This greater attention to item content should, therefore, reduce both response set and satisficing; while also balancing the impact of yea- or nay-saying. Nunnally (1978) proposed that a balance of positively- and negatively-worded items would virtually eliminate such behaviors. However, an increasing number of researchers are recommending that this practice should be undertaken with caution.

Several studies have found that item orientation can potentially confound factor structure (Campbell & Grissom, 1979; Deemer & Minke, 1999; Eggers, 2000; Johnson

& Osborn, 2000) and may often result in a separate factor for the negatively-worded items (Anderson, Anderson, & Jenson, 1979; Ibrhim, 2001; Magazine, Williams, & Williams, 1996; McInerney, McInerney, & Roche, 1994; Motl, Conroy, & Horan, 2000). The presence of such a factor is disturbing in that it is "irrelevant to the trait being measured" (Ibrahim, 2001, p. 498). However, the debate over the cause of the negative factor continues. To date, researchers have attributed the cause to "careless responses (Schmitt & Stults, 1985), insufficient cognitive ability (Cordery & Sevastos, 1993), impaired response accuracy as a result of the negatively-worded items (Schriesheim, Eisenbach, & Hill, 1991; Schriesheim & Hill, 1981), and the actual measurement of a different construct (Pilotte & Gable, 1990)" (Magazine et al., 1996, p. 247).

Lustig (1963) investigated the relationship between response styles of yes-saying, nay-saying, and in-between-saying, and perceptual aspects of personality for a sample of 220 high school and college students. Lustig developed an instrument to categorize participant response styles and then compared the style groups on several personality traits. Findings indicated that, yea-sayers viewed the world as more positive and friendly, nea-sayers perceived the world as more negative and unfriendly, and in-between sayers saw the world as more ambivalent and uncertain. The problem of yea-saying was more common in that twice as many participants fell into that category.

A respondent characteristic that has been explored in several studies is age. Melnick and Gable (1990) examined responses to positively- and negatively-worded items on the Parent Attitudes Toward School Effectiveness survey (Gable, Murphy, Hall, & Clark, 1986) completed by 3,328 parents. These researches found that age,

education level, interest in topic, and readability of the instrument all impacted responses to the mixed stem items. Eggers (2000) found that senior adults responded differently to the negatively-worded items, in that the items did not load on the anticipated factors; however, the data from the mid-life group fit the expected model. Benson and Hocevar (1985) detected an age effect in that middle school students were less likely to indicate agreement by disagreeing to a negatively-worded item. This finding, however, could be attributed to reading ability. Marsh (1984) found that pre-adolescent children often responded to negatively-worded items inappropriately and detected a relationship between the inappropriate responses and reading ability that was independent of age. Therefore, the effect could have resulted from either age, educational level, or a combination of the two.

Of course, no discussion of responses to survey items would be complete without discussing score reliability and validity. Barnette (2000) found a substantial decrease in reliability for the negatively-worded items on an attitude toward year-round schooling survey administered to high school students. Johnson and Osborn (2000) found an increase in reliability on responses to a theoretical orientation scale when yea-sayers and nay-sayers were dropped from the analysis. Sandoval and Lambert (1978) found that adding positively-worded items to a teacher rating of a hyperactivity instrument composed of negatively-worded items increased both score reliability and validity. Somewhat in contrast, Schriesheim and Hill (1981) found that negatively-worded items increased validity but had no impact on the reliability of scores from undergraduates providing responses to a presented scenario.

Wright and Master (1982) concluded that the differently worded items did not

provide consistent information. Congruent with this conclusion, using generalizability theory, Chang (1995) established that negatively-worded items are not fully equivalent to their positively-worded counterparts. Moreover, Johanson, Gips, and Rich (1993) found that the negatively-worded items may result in a lack of information (i.e, non-response). They concluded that participants may be following the social norm that if you do not have something nice to say then do not say anything at all; therefore, less favorable responses may be omitted more often.

However, not all studies have found differences in responses to positively- and negatively-worded items. Marsh (1986) employed a construct-validity approach to data from three studies and found no reason to separate the items into positive and negative subscales. Similarly, Bergstrom and Lunz (1998) utilized item response theory and concluded that the positively- and negatively-worded items appeared to measure the same construct. On a dichotomous scale administered to middle school students, Williams, Bush, Par, Malone, and Jessup (2001) noted that responses to negatively-worded items were more strongly related to the criterion than were responses to positively-worded items.

Therefore, discrepancies exist among the findings of studies exploring the impact of positively- and negatively-worded items. Primary causes for the incongruity include differences in survey content, methodology, participants, and criteria for analyzing the impact. The inconsistency could also be attributed to the comparison of responses to positively- and negatively-worded items that are not exact semantic opposites. Studies focusing on minimizing non-attending behaviors will undoubtedly result in a variety of findings due to the degree of nonattentiveness cultivated by differences in both content

and participants. One participant characteristics that consistently emerges, however, is age. It is yet to be determined, however, whether the age of the participant is associated with non-attending because the participants are careless in their responses or are confused by the negative wording.

While age has been identified as a potential characteristic, what other characteristics can be identified? Weems, Onwuegbuzie, Eggers, and Schrieber (2001) found that participants with the greatest differential in responses between positively- and negatively-worded items on a measure of research anxiety tended to have negative self-perceptions about their academic competence, to have the highest levels of hope associated with pathways, not to have tendencies towards cooperative learning, and not to be self-oriented perfectionists. However, as noted by the researchers, much more work is needed in this area. Therefore, the purpose of this study is to replicate and to extend the study of Weems et al.'s study by investigating the characteristics of respondents who respond differently to positively- and negatively-worded items. The data for this study came from two samples gathered in larger research efforts.

STUDY 1

Participants

The sample comprised 185 students enrolled in Spanish (63.93%), French (25.57%), German (7.76%), or Japanese (2.74%) introductory-level courses at a mid-southern university. The ages of the respondents ranged from 18 to 71 (M = 22.78, SD = 6.92), with 33.2% being male. The mean grade point average (gpa) was 3.05 (SD = 0.59. Participation was voluntary. A series of Kruskal-Wallis one-way analyses of variance revealed no differences ($p > .05$) across the courses with respect to the three

measures of foreign language anxiety, foreign language achievement, gpa, age, and measures of self-perception, study skills, and locus of control; thus, responses of all participants were combined.

Instruments and Procedure

A battery of instruments was used in the study, namely: the Input Anxiety Scale (IAS), the Processing Anxiety Scale (PAS), the Output Anxiety Scale (OAS), the Self-Perception Profile for College Students (SPPCS), the Academic Locus of Control Scale (ALC), the Study Habits Inventory (SHI), and the Background Demographic Form (BDF). Participants were given the questionnaire packet containing these instruments during the fourth week of the semester. They were instructed to complete the battery of instruments at home and to return it within two weeks.

The three anxiety scales (i.e., the Input Anxiety Scale, the Processing Anxiety Scale, the Output Anxiety Scale) were developed by MacIntyre and Gardner (1994). Each scale contains six 5-point Likert-format items (i.e., 1 = *strongly agree*, 2 = *agree*, 3 = *neutral*, 4 = *disagree*, 5 = *strongly disagree*) that assess how anxious students feel at the input, processing, and output stages of the foreign language learning process. For each scale, three items are positively worded and three items are negatively worded. All negative items were key-reversed before scoring, such that high scores on any of these scales represent high levels of anxiety at the corresponding stage. Sample items for the Input Anxiety Scale include, "I get flustered unless French/German/Spanish is spoken very slowly and deliberately" and "I get upset when I read in French/German/Spanish because I must read things again and again." Sample items for the Processing Anxiety Scale include, "I am anxious with French/German/Spanish

because, no matter how hard I try, I have trouble understanding it" and "I feel anxious if French/German/Spanish class seems disorganized." Finally, sample items for the Output Anxiety Scale include, "I may know the proper French/German/Spanish expression but when I am nervous it just won't come out" and "When I become anxious during a French/German/Spanish test, I cannot remember anything I studied." For the present inquiry, the scores pertaining to the Input Anxiety Scale, the Processing Anxiety Scale, and the Output Anxiety Scale had classical theory alpha reliability coefficients of .70 (95% confidence interval [CI] = .63, .76), .73 (95% CI = .66, .79), and .76 (95% CI = .70, .81), respectively.

The SPPCS (Neemann & Harter, 1986) is a 54-item scale consisting of 13 subscales (i.e., perceived creativity, perceived intellectual ability, perceived scholastic competence, perceived job competence, perceived athletic competence, perceived appearance, perceived romantic relationships, perceived social acceptance, perceived close friendships, perceived parent relationships, perceived humor, perceived morality, and perceived global self-worth). In order to ensure model parsimony, only the perceived intellectual ability, perceived scholastic competence, perceived social acceptance, and perceived global self-worth subscales of the SPPCS were analyzed in this study. For the present study, the classical theory alpha reliability coefficients pertaining to the selected subscales were as follows: perceived intellectual ability (.79; 95% CI = .74, .84), perceived scholastic competence (.71; 95% CI = .64, .77), perceived social acceptance (.83; 95% CI = .79, .87), and perceived global self-worth (.87; 95% CI = .84, .90).

The ALC, developed by Trice (1985), has 28 true-false items related to personal

control over academic outcomes. Scores range from 1 (strongly internal locus) to 28 (strongly external locus). For the current investigation, the classical theory alpha reliability coefficient for the ALC was .73 (95% CI = .67, .78).

The SHI, developed by Jones and Slate (1992), consists of 63 true-false items designed to assess the typical study behaviors of college students. Thirty items describe effective study behaviors, and 33 items specify ineffective study behaviors. The latter items are key-reversed such that total scale scores range from 0 to 63, with high scores indicating good study skills. For the present research, the classical theory alpha reliability coefficient for the SHI was .88 (95% CI = .85, .90).

The BDF, developed specifically for this study, extracted relevant information such as age, gender, and students' expectations for their overall average in their current language course. This latter variable was measured on a 100-point scale.

Finally, foreign language achievement was measured using students' course averages. The course averages were measured on a 100-point scale. This global measure was selected instead of an isolated measure of specific skills in order to maximize the external validity (i.e., generalizability) of the findings. In order to adjust for differences in teacher characteristics (e.g., effectiveness, experience, motivation, and testing and scoring standards), standardized course averages were used instead of raw averages. Standardized course averages (i.e., *z*-scores) were computed for each student by subtracting the average achievement score of the foreign language class to which the student belonged from the student's course average, and then dividing by the class standard deviation.

Analysis

A series of dependent t-tests was used to compare scores from the positively-worded items and scores from the negatively-worded items for each of the three anxiety scales (i.e., IAS, PAS, OAS). Bonferroni's adjustment was used to maintain an overall 5% level of significance. For the major analysis, a canonical correlation analysis was used to determine the characteristics of students who had the greatest differential in responses between the positively- and negatively-worded items on the three measures of anxiety. Canonical correlation analysis is a technique used to assess the relationship between two sets of variables when each set contains at least two variables (Cliff & Krus, 1976; Darlington, Weinberg, & Walberg, 1973; Thompson, 1980, 1984). The absolute difference between scores on the positively-worded items ($n = 6$) and the negatively-worded items ($n = 6$) for the IAS, PAS, and OAS served as the three dependent variables (i.e., dependent multivariate set of variables), whereas the independent set of variables consisted of gender, age, grade point average, whether the foreign language was required, students' expectations for their overall average in their current language course, number of high school foreign language courses taken, number of university foreign language courses taken, overall course achievement in the foreign language course, study habits, locus of control, perceived intellectual ability, perceived scholastic competence, perceived social acceptance, and perceived global self-worth. The number of canonical functions (i.e., factors) that can be generated for a given dataset is equal to the number of variables in the smaller of the two variable sets. Because three anxiety measures were involved, three canonical functions were generated.

Results

The series of dependent t-tests revealed that for the IAS, scores from the positively-worded items (M = 3.11, SD = 0.68) were statistically significantly (t = 3.72, p < .001) higher than scores from the negatively-worded items (M = 2.92, SD = 0.83). The effect size associated with this difference, as measured by Cohen's (1988) d, was 0.25, which could be considered as being indicative of a small effect. Conversely, for the PAS, scores from the negatively-worded items (M = 3.04, SD = 0.75) were statistically significantly (t = 5.12, p < .0001) higher than scores from the positively-worded items (M = 2.76, SD = 0.81). The associated effect size of .36 was moderate (Cohen, 1988). Finally, for the OAS, no statistically significant difference (t = 0.39, p > .05) emerged between scores from the positively-worded items (M = 3.18, SD = 0.76) and scores from the negatively-worded items (M = 3.16, SD = 0.82).

A series of score reliability coefficients was computed on the positively- and negatively-worded items pertaining to the three anxiety scales. For each scale, the score reliability indices for the positively-worded items were very different than those for the negatively-worded items. Specifically, for the IAS, the classical theory alpha reliability coefficient was .31 (95% CI = .12, .47) for the positively-worded items and .74 (95% CI = .67, .80) for the negatively-worded items. For the PAS, the score reliability coefficient was .71 (95% CI = .63, .78) for the positively-worded items and .55 (95% CI = .42, .65) for the negatively-worded items. Finally, for the OAS, the score reliability coefficient was .59 (95% CI = .48, .68) for the positively-worded items and .70 (95% CI = .62, .77) for the negatively-worded items.

The strength of the relationship between the two sets of variables was assessed by examining the magnitude of the canonical correlation coefficients. These

coefficients indicate the degree of relationship between the weighted anxiety variables and the weighted set of independent variables. In addition, the significance of the canonical roots was tested via the F-statistic based on Rao's approximation (Rao, 1952). (The full correlation matrix that generated the canonical correlation analysis, although not presented due to space constraints, can be obtained from the authors.)

The canonical analysis revealed that all three canonical correlations combined were statistically significant (F [142, 499.13] = 1.42, $p < .05$). Further, when the first canonical root was excluded, the remaining two canonical roots were not statistically significant. Similarly, when the first two canonical roots were excluded, the remaining canonical root was not statistically significant. Together, these results suggested that only the first ($R_{c1} = .47$) canonical function was statistically significant, and represented a moderate effect size, contributing 21.7% (i.e., R_{c1}^2) to the shared variance. Both the second ($R_{c1}^2 = 5.1\%$) and third ($R_{c1}^2 = 3.7\%$) canonical functions did not contribute much to the shared variance. Thus, only the first canonical correlation was interpreted.

Data pertaining to the first canonical root are presented in Table 1. This table provides both the standardized function coefficient and the structure coefficient pertaining to the first canonical correlation. An examination of the standardized canonical function coefficients revealed that, using a cutoff correlation of 0.3 recommended by Lambert and Durand (1975) as an acceptable minimum loading value, positive-negative differentials for both the IAS and PAS made important contributions to the anxiety composite, with the differential corresponding to the PAS making the greatest contribution. Interestingly, the IAS and PAS differentials were inversely related–consistent with the dependent t-test results. With respect to the

independent variables set, locus of control and students' expectations for their overall average in their current language course made important contributions to the composite set.

Similarly, the structure coefficients pertaining to the first canonical correlation revealed that positive-negative differentials for both the IAS and PAS made important contributions to the model, again in an inverse manner. Also, the following seven independent variables made important contributions: students' expectations for their overall average in their current language course, study habits, locus of control, perceived intellectual ability, perceived scholastic competence, perceived social acceptance, and perceived global self-worth.

Interestingly, from Table 1, it can be seen from the square of the structural coefficients that locus of control was the best predictor of positive-negative differentials, followed by perceived global self-worth, study habits, perceived social acceptance, perceived intellectual ability, perceived scholastic competence, and students' expectations for their overall average in their current language course, respectively.

Insert Table 1 about here

STUDY 2

Participants

The archival data set used for the second study represented a convenience sample of 86 participants recruited from a client database of a college disability services center. Participation was voluntary, anonymous, and informed consent was

obtained. The ages of the participants ranged from 18 to 51 (M = 27.10, SD = 10.20), with 82% of the enrolled in undergraduate programs and 18% enrolled in graduate programs. The majority of the sample was female (58.8%), Caucasian (93%), and had never been married (80%). Cumulative undergraduate GPA of participants was reported as 8.33% below 2.0, 51.19% from 2.0 to 2.9, and 40.48% 3.0 or higher.

Instrument

The data were collected for the purpose of examining the relationship between an individual's sense of coherence and dysfunctional career thoughts. The Sense of Coherence Scale (SOCS, Antonovsky, 1987) is a 29 item scale measuring sense of coherence with scales addressing comprehensibility, manageability (MA), and meaningfulness (ME). The SOCS utilizes seven-point bipolar adjective scales with unique adjectives provided on for virtually all items. The comprehensibility scale was not utilized in this study because only two items were negatively-worded. Sample items from the MA scale include "Has it happened that people whom you counted on disappoint you? (Adjectives: never happened and always happened)" and " What best describes how you see life? (Adjectives: one can always find a solution to painful things in life and there is no solution to painful things in life)." Sample items from the ME scale include "Life is: (Adjectives: full of interest and completely routine)" and " When you think about life, you very often: (Adjectives: feel how good it is to be alive and ask yourself why you exist at all)." For the present study, responses for scales were averaged with higher scores indicating a stronger sense of coherence. Several researchers have examined the criterion validity of scores obtained by the SOCS. Antonovsky (1993) reviewed 42 studies and found positive correlations between SOCS

and (a) health and well being, (b) self esteem, (c) social skills, and (d) social support; and negative correlations with anxiety and perceived stressors. For the present study, the classical theory alpha reliability coefficient for the MA scale was .83 (95% CI = .77, .88) and for the ME scale was .85 (95% CI = .79, .89).

Procedures

To identify characteristics of participants who responded differently to negatively-worded and positively-worded items, separate scale scores were calculated by item wording for both the MA and ME indices. The absolute differences between the scores were then used as the dependent variable in the regression analysis. Because of the mixed results concerning item orientation, scale means and reliabilities were compared before regressing the difference scores.

Results

Reliabilities for the positively- and negatively-worded items within the two scales were not compared for this study due to the asymmetric division and the small number of items once the scales were divided and the impact of scale length on scale reliability. After deleting one case for an out-of-range value, scores from the positively-worded MA scale (M = 4.55, SD = 1.15) and the negatively-worded MA scale (M = 5.02, SD = 1.14) were compared with a dependent t-test. Results indicated that the scores from the two scales differed significantly ($t(84) = 5.128$, $p < .001$), in favor of the positively-worded scale with a moderate effect size (d = 0.55). A dependent t-test also was used to compare scores from the positively-worded ME scale (M = 5.25, SD = 1.27) and the negatively-worded ME scale (M = 5.32, SD = 1.04). No statistically significant differences emerged between responses to the positively- and negatively-worded items

on the ME scale ($t(84) = 0.771$, $p < .443$).

A standard multiple regression was performed, with the absolute value of the difference scores on the MA scale serving as the dependent variable and gender, age, marital status (never married vs. other), and GPA (below 2.0, 2.0 to 2.9, and 3.0 or higher) entered as independent variables. The assumptions for regression seemed reasonable. Using a $p < .001$ criterion for Mahalanobis distance revealed no outliers (Maximum = 17.72); and using Cook's d did not highlight any influential observations (Maximum = 0.13). The normal probability plot of the standardized residuals suggested normality, and multicollinearity did not appear to be a problem; additionally the maximum variance inflation factor was 2.35. Missing data forced the omission of 5 cases resulting in a final sample size of $n = 80$. The multiple correlation coefficient, R, however, was not statistically significantly different from zero, $F(4, 75) = 0.69$, $p > 0.05$.

A standard multiple regression also was performed, with the absolute value of the difference scores on the ME scale serving as the dependent variable and gender, age, marital status (never married vs. other), and GPA (below 2.0, 2.0 to 2.9, and 3.0 or higher) entered as independent variables. The assumptions for regression seemed reasonable. Using a $u < .001$ criterion for Mahalanobis distance, no outliers were detected (Maximum = 17.72); and using Cook's d indicated no influential observations (Maximum = 0.09). The normal probability plot of the standardized residuals suggested normality and multicollinearity did not appear to be a problem; additionally the maximum variance inflation factor was 2.35. Missing data forced the omission of 5 cases resulting in a final sample size of 80. Again, the multiple correlation coefficient, R, again was not statistically significantly different from zero, $F(4, 75) = 1.16$, $p > 0.05$.

DISCUSSION

Many instruments, especially Likert-type scales, contain both positively- and negatively-worded items within the same scale (i.e., mixed item format). A major reason for this practice appears to be to discourage response sets from emerging. Using this format also helps the analyst to detect response sets that occur in data sets, and thus eliminate them from subsequent analyses. However, some psychometricians (e.g., Barnette, 2000) seriously question the use of mixed item formats, positing that positively- and negatively-worded items within a scale are not measuring the same underlying trait.

Recently, evidence has suggested that certain individuals are more predisposed to providing differential response patterns when responding to a mixed item format scale (Weems et al., 2001). However, to date, only a few characteristics of these differential-responding individuals have been identified. Thus, the purpose of this present study was to extend this line of research. Specifically, two studies were conducted in which responses to several scales that utilize mixed item formats were analyzed.

Participants in Study 1 responded differently to each of the three measures of foreign language anxiety. Whereas responses to the positively- and negatively-worded items of the Output Anxiety Scale were very consistent, this was not the case for either the Input Anxiety Scale or the Output Anxiety Scale. Moreover, these two latter scales evoked opposite sets of response patterns. Specifically, for the IAS scale, scores pertaining to the positively-worded items were moderately higher than were scores from the negatively-worded items, whereas the converse was true for the PAS. These

findings combined suggest that on two of the three anxiety scales, the positively- and negatively-worded items possibly measured different constructs.

Interestingly, certain characteristics of the participants made them more likely to generate differential patterns of responses to the positively- and negatively-worded items. In particular, undergraduate students who had the highest expectations for their overall average in their current language course, the most effective study habits, the most internal locus of control, and the highest levels of perceived intellectual ability, perceived scholastic competence, perceived social acceptance, and perceived global self-worth tended to have greatest differential in responses between positively- and negatively-worded items on the Input Anxiety Scale and the least differential on the Processing Anxiety Scale. Simply put, students with the most positive orientation with respect to course expectations, study habits, locus of control, and self-perceptions tended to provide more extreme responses to the positive items on the Input Anxiety Scale and to provide more extreme responses to the negative items on the Processing Anxiety Scale. The reverse was true for students with the least positive orientations.

Weems et al. (2001) found that graduate students with the greatest differential in responses between positively- and negatively-worded items tended to have certain traits. Thus, the results in Study 1 are consistent with this finding. Both of these studies suggest that strongly disagreeing to a positively-worded item is not equivalent to strongly agreeing to a negatively-worded item. Thus, using a mixed item formats may represent an important threat to the content- and construct-related validity of a scale, especially for certain individuals. In fact, because the positively- and negatively-worded items on two of the scales in Study 1 induced very different responses, it appears that

use of multidimensional scales that contain mixed item formats within the same study yield even more complex differential response patterns.

However, the most disturbing aspect of the data in Study 1 was the fact that across the three scales, the reliability indices pertaining to scores on the positively- and negatively-worded items were very different. Whereas for the PAS, the score reliability coefficient was 29.1% higher for the positively-worded items, for the IAS and OAS, the score reliability coefficient was 138.7% and 18..6% higher for the negatively-worded items, respectively. Thus, although the IAS, PAS, and OAS scales all yielded scores with adequate reliability coefficients (i.e., > .70; Nunnally & Bernstein, 1994), it is likely that these indices were attenuated by the use of mixed item formats. That is, for the PAS, use of positively-worded items severely reduced score reliability. Conversely, for both the IAS and OAS, use of negatively-worded items severely reduced score reliability.

Using the Spearman-Brown prophecy (Crocker & Algina, 1986) suggests that if the three positively-worded items on the PAS had been replaced by three parallel negatively-worded items, the score reliability for the total scale would increase from .70 to .82–which represents a 17.1% increase. Similarly, if the three negatively-worded items on the IAS had been replaced by three parallel positively-worded items, the score reliability for the total scale would increase from .73 to .84 (a 15.1% increase). Also, if the three negatively-worded items on the OAS had been replaced by three parallel positively-worded items, the score reliability for the total scale would increase from .76 to .86 (a 13.2% increase). This is extremely compelling evidence against the use of mixed item formats.

The results from Study 2 did not uncover similar findings. The only similarity was a significant difference between responses to the positively- and negatively-worded items on the MA scale. Multiple regression on neither the MA or ME scale uncovered respondent characteristics that were associated with different response patterns. Study 2, however, was more limited by a minimum of participant characteristics available in the archival data set and the small sample size greatly attenuated the statistical power for the regression analysis.

However, a key distinction of the second study was the type of response options offered. Rather than responding to a series of Likert-type items, participants responded using bipolar adjectives. The adjectives offered changed with virtually every item, and the orientation of the items also were randomly altered between positively- and negatively-worded items. Thus, the participants may have been responding more carefully due to a potential novelty effect and, therefore, different response patterns to reverse-coded items due to carelessness and/or response set were virtually eliminated. The adjectives provided in this study also differed from the Likert-type scale in that participants were not asked to "disagree." Therefore, the response options did not facilitate either "yea-saying" or "nay-saying" by participants.

CONCLUSION

The first study yielded extremely compelling evidence against the use of mixed item format. In this investigation, responses to positively- and negatively-worded items were significantly different. Indeed, in the case of the IAS, the response options appeared to facilitate "nay-saying," whereas with respect to the PAS, "yea-saying" seemed to prevail among the participants. Further, the extent of the response

differential was a function of various individual characteristics. Moreover, across all three anxiety scales, the item format that yielded the highest mean scores also yielded the significantly lowest score reliability coefficient. These results suggest that writing an item that is inappropriately worded (i.e., either positively- or negatively-worded) induces artificially extreme responses, which, in turn, attenuates the score reliability of the scale. As was the case in the study of Weems et al. (2001), the first set of results suggest strongly that use of positively- and negatively-worded items within the same scale may seriously threaten both score reliability and score validity. As admonished by Weems et al., use of mixed item formats should be undertaken with extreme caution.

In the second study, the significant difference between the positively- and negatively-worded items on the MA scale lends further support to suggest that the item wordings possibly measured different constructs. The second study also suggests that more research is needed to ascertain whether responses to Likert-formated scales are more adversely affected by mixed item formats than are other response options such as bipolar adjectives or if the lack of findings in this study were the result of low statistical power. In light of the findings from the present study, and similar findings in the literature reviewed, a replication of Lustig's (1963) study to explore further the relationship between differential response patterns and personality characteristics by utilizing his instrument designed to identify acquiescence.

Together, the two studies suggest that either (a) respondents process positively-worded items differently than negatively-worded items or (b) respondents do not read the negatively worded items as carefully as they do positively-worded items, and simply agree to the negatively-worded items instead of disagreeing due to a response set.

Second, several characteristics were identified pertaining to those who tend to have the largest absolute discrepancies in responses between the two sets of items.

Indeed, bearing in mind that such scale formats can attenuate reliability estimates (Weems & Onwuegbuzie, 2001), when such formats are used, the resultant score reliability should be scrutinized. Further, researchers should refrain from comparing subscales that have different item wording formats. In particular, it is likely that comparing a subscale with positively- and negatively-worded items to another subscale whose wording is in the same direction will yield misleading results arising from different response patterns that have little to do with the actual constructs of interest, but, rather reflect acquiescing.

REFERENCES

Anderson, J. P., Anderson, P. A., & Jenson, A. D. (1979). The measurement of Nonverbal Immediacy. Paper presented at the annual meeting of the Eastern Communication Association, Philadelphia, PA. (ERIC Document Reproduction Service No. ED 203 409)

Antonovsky, A. (1987). Unraveling the mystery of health. San Fransisco: Jossey-Bass.

Antonovsky, A. (1993). The structure and properties of the Sense of Coherence Scale. Social Science and Medicine, 36, 725-733.

Barnette, J. J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: If you feel the need, there is a better alternative to using those negatively worded stems. Educational and Psychological Measurement, 60(3), 361-370.

Benson, J., & Hocevar, D. (1985). The impact of item phrasing on the validity of attitude scales for elementary school children. Journal of Educational Measurement, 22, 231-240.

Bergstrom, B. A, & Lunz, M. E. (1998). Rating scale analysis: Gauging the impact of positively and negatively worded items. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA. (ERIC Document Reproduction Service No. ED 423 289)

Campbell, N. J., & Grissom, S. (1979). Influence of item direction on student responses in attitude assessment. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. ED 170 366)

Chang, L. (1995). Connotatively consistent and reversed connotatively inconsistent items are not fully equivalent: Generalizability study. Educational and Psychological Measurement, 55(6), 991-997.

Cliff, N., & Krus, D. J. (1976). Interpretation of canonical analyses: Rotated vs. unrotated solutions. Psychometrica, 41, 35-42.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. New York: John Wiley.

Couch, A., & Keniston, K. (1960). Yea sayers and nay sayers: Agreeing response set as a personality variable. Journal of Abnormal and Social Psychology, 60, 151-174.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Orlando, FL: Holt, Rinehart, and Winston.

Cronbach, L. J. (1946). Response sets and test validity. Educational and Psychological Measurement, 6, 475-494.

Cronbach, L. J. (1950). Further evidence on response sets and test design. Educational and Psychological Measurement, 10, 3-31.

Daly, J. A., & Miller, M. D. (1975). The empirical development of an instrument to measure writing apprehension. Research in the Teaching of English, 9, 242-249.

Darlington, R. B., Weinberg, S. L., & Walberg, H. J. (1973). Canonical variate analysis and related techniques. Review of Educational Research, 42, 131-143.

Deemer, S.A., & Minke, K. M. (1999). An investigation of the factor structure of the teacher efficacy scale. Journal of Educational Research, 93(1), 3-10.

Dunn, R., Dunn, K., & Price, G. (1991). Productivity Environmental Preference Survey. Lawrence, KS: Price Systems.

Eggers, S. J. (2000). Spirituality in mid-life and late adulthood (Doctoral dissertation, The University of Memphis, 2000). Dissertation Abstracts International, 61,(3A), 878.

Gable, R., Murphy, C., Hall, C., & Clark, A. (1986, April). The development of the pilot form of the Parent Attitudes Toward School Effectiveness (PATSE) questionnaire. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Hewitt, P.L., & Flett, G.L. (1991). Perfectionism in self and social contexts: Conceptualization, assessment, and association with psychopathology. Journal of Personality and Social Psychology, 60, 456-470.

Ibrahim, A. M. (2001). Differential responding to positive and negative items: The case of a negative item in a questionnaire for course and faculty evaluation. Psychological Reports, 88, 497-500.

Jones, C. H., & Slate, J. R. (1992). Technical Manual for the Study Habits Inventory. Unpublished manuscript, Arkansas State University, AR: State University.

Johnson, G. A., Gips, C. J., & Fich, C. E. (1993). If you can't say something nice: A variation on the social desirability of response set. Evaluation Review, 17(1), 116-122.

Johnson, G. A., & Osborn, C. J. (2000). Acquiescence as differential person functioning. Paper presented at the annual meeting of the American Education

Research Association, New Orlean, LA. (ERIC Document Reproduction Service No. ED 441 022)

Jones, C.H., & Slate, J.R. (1992). Technical manual for the Study Habits Inventory. Unpublished manuscript, Arkansas State University, Jonesboro, AR.

Krosnick, J. A., Narayan, S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. In Braverman, M. T., & Slater, J. K. (Eds.), Advances in survey research (pp. 29-44). San Francisco: Sage.

Lambert, Z.V., & Durand, R.M. (1975). Some precautions in using canonical analysis. Journal of Market Research, XII, 468-475.

Lustig, P. (1963). Response style and personality. Unpublished doctoral dissertation, New York University.

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. Language Learning, 44, 283-305.

Magazine, S. L., Williams, L. J., & Williams, M. L. (1996). A confirmatory factor analysis examination of reverse coding effects in Meyer and Allen's Affective and Continuance Commitment Scales. Educational and Psychological Measurement, 56(2), 241-250.

Marsh, H. W. (1984). The bias of negatively worded items in rating scales for preadolescent children: A cognitive-developmental phenomenon. (ERIC Document Reproduction Service No. ED 242 772)

Marsh, H. W. (1986). Multidimensional self concepts: Do positively and negatively worded items measure substatively different components of self. (ERIC Document Reproduction Service No. ED 278 716)

McInerney, V., McInerney, D., & Roche, L. (1994). Definitely not just another computer anxiety instrument: The development and validation of CALM: Computer Anxiety and Learning Measure. Paper presented at the annual Stress and Anxiety Research Conference, Madrid, Spain. (ERIC Document Reproduction Service No. ED 386 161).

Melnick, S. A., & Gable, R. K. The use of negative item stems: A cautionary note. Educational Research Quarterly, 14(3), 31-36.

Motl, R. W., Conroy, D. E., & Horan, P. M. (2000). The Social Physique Anxiety Scale: An example of the potential consequence of negatively worded items in factorial validity studies, 1(4), 327-345.

Myers, R.H. (1986). Classical and modern regression with applications. Boston, MA: Duxbury Press.

Neemann, J., & Harter, S. (1986). Manual for the Self-perception Profile for College Students. Unpublished manuscript, University of Denver, CO: University of Denver.

Nunnally, J. C. (1978). Psychometric Theory (2nd ed.). New York: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). New York: McGraw-Hill.

Onwuegbuzie, A.J. (1996). Development of the Research Anxiety Rating Scale. Unpublished manuscript, University of Central Arkansas.

Onwuegbuzie, A.J., & Daniel, L.G. (in press). Typology of analytical and interpretational errors in quantitative and qualitative educational research. Current Issues in Education.

Onwuegbuzie, A.J., Slate, J., Paterson, F., Watson, M., & Schwartz, R. (2000). Factors associated with underachievement in educational research courses. Research in the Schools, 7, 53-65.

Rao, C. R. (1952). Advanced statistical methods in biometric research. New York: Wiley.

Sandoval, J., & Lambert, N. M. (1978). Reliability and validity of teacher rating procedures in the assessment of hyperactivity as a function of rating scale format. Paper presented at the annual meeting of the American Educational Research Association, Toronto, Canada. (ERIC Document Reproduction Service No. ED 160 614)

Schmeck, R. R., Geisler-Brenstein, E, & Cercy, S. P. (1991). Self-concept and learning: The revised inventory of learning processes. Educational Psychology, 11(3), 343-362.

Schriesheim, C. A., & Hill, K. D. (1981). Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. Educational and Psychological Measurement, 41, 1101-1114.

Sen, A.K., & Srivastava, M. (1990). Regression analysis: Theory, methods, and applications. New York: Springer-Verlag.

Shapiro, S.S., & Wilk, M.B. (1965). An analysis of variance test, for normality and complete samples. Biometrika, 52, 592-611.

Shapiro, S.S., Wilk, M.B., & Chen, H.J. (1968). A comparative study of various tests for normality. Journal of the American Statistical Association, 63, 1343-1372.

Snyder, C.R., Harris, C., Anderson, J.R., Holleran, S.A., Irving, L.M., Sigmon, S.T., Yoshinobu, L. Gibb, J., Langelle, C., & Harney, P. (1991). The will and the ways: Development and validation of an individual-differences measure of hope. Journal of Personality and Social Psychology, 60, 570-585.

Thompson, B. (1980, April). Canonical correlation: Recent extensions for modelling educational processes. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Thompson, B. (1984). Canonical correlation analysis: Uses and interpretations. Newbury Park, CA: Sage Publications. (ERIC Document Reproduction Service No. ED 199 269)

Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. Educational and Psychological Measurement, 55, 525-534.

Thompson, B. (1998, April). Five methodology errors in educational research: The pantheon of statistical significance and other faux pas. Invited address presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Thompson, B., & Borrello, G. M. (1985). The importance of structure coefficients in regression research. Educational and Psychological Measurement, 45, 203-209.

Trice, A. (1985). An academic locus of control scale for college students. Perceptual and Motor Skills, 61, 1043-1046.

Weems, G. H. & Onwuegbuzie, A. J. (2001). The impact of midpoint responses and reverse coding on survey data. Measurement and Evaluation in Counseling and Development, 34(3), 166-176.

Weems, G. H., Onwuegbuzie, A. J., Eggers, S. J., & Schrieber, J. B. (2001). Characteristics of respondents with different response patterns to positively-worded and negatively-worded items on rating scales. Paper presented at the annual meeting of the Mid-South Educational Research Association.

Williams, R. L., Bush, V. J., Park, S. H., Malone, Y., & Jessup, K. (2001). Work ethic scale for middle school students. (ERIC Document Reproduction Service No. ED 452 257)

Wright, B. D., & Masters, G. N. (1982). Rating scale analysis. Chicago: Mesa Press.

Table 1

Canonical Solution for First Function

Variable	Standardized Coefficient	Structure Coefficient	Structure Coefficient^2 (%)
Criterion Set:			
Input Anxiety Scale	0.63	0.62	38.44
Processing Anxiety Scale	-0.81	-0.70	49.00
Output Anxiety Scale	0.27	0.14	1.96
Predictor Set:			
gender	-0.03	-0.05	0.25
age	-0.16	-0.10	1.00
grade point average	-0.24	-0.10	1.00
foreign language required	0.13	0.16	2.56
students' expectations	0.45*	0.47*	22.09
number of high school language courses	0.15	0.12	1.44
number of university language courses	-0.19	-0.28	7.84
overall language course achievement	-0.26	-0.03	0.09
study habits	0.19	0.56*	31.36
locus of control	-0.57*	-0.73*	53.29
perceived intellectual ability	0.10	0.52*	27.04
perceived scholastic competence	-0.01	0.51*	26.01
perceived social acceptance	0.02	0.53*	28.09
perceived global self-worth	0.12	0.57*	32.49

* loadings with large effect sizes (Lambert & Durand, 1975)

TM034646

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Profiles of Respondents who Respond Inconsistently to Positively - and Negatively - Worded Items on Rating Scales

Author(s): Gail H. Weems, Anthony J Onwuegbuzie, & Daniel Lustig

Corporate Source:	Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☑ Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐ Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐ Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature: Gail H. Weems	Printed Name/Position/Title: Gail Weems Assistant Professor	
Organization/Address: University of Memphis Ball Hall 100 Memphis, TN 38152	Telephone: 901 678 1770	FAX: 901.678.5440
	E-Mail Address: ghweems@memphis.edu	Date: Nov. 7, 2002

(Over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20742-5701
ATTN: ACQUISITIONS

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: info@ericfac.piccard.csc.com
WWW: http://ericfacility.org

EFF-088 (Rev. 2/2001)